3К

УДК 004.89:004.93

Н.С. Клименко
Институт проблем искусственного интеллекта
МОН Украины и НАН Украины, г. Донецк
Украина, 83048, г. Донецк, ул. Артема, 118-б

Разработка структуры текстонезависимой системы идентификации диктора

M.S. Klymenko
Institute of Artificial Intelligence
MES of Ukraine and MAS of Ukraine, c. Donetsk
Ukraine, 83048, c. Donetsk, Artema st., 118-b

Development of Structure for Text-Independent Speaker Identification System

М.С. Клименко
Інститут проблем штучного інтелекту
МОН України і НАН України, м. Донецьк
Україна, 83048, м. Донецьк, вул. Артема, 118-б

Розробка структури текстонезалежної системи ідентифікації диктора

В статье рассмотрены основные технологии, используемые при создании систем идентификации диктора, и трудности, с которыми сталкиваются их разработчики. Предложена структура системы текстонезависимой идентификации диктора, использующая автоматическую дикторонезависимую сегментацию речевого сигнала с одновременной классификацией сегментов. Такой подход повышает точность модели диктора и нивелирует разногласие между обучающим и распознаваемым контекстом.
Ключевые слова: идентификация личности по голосу, широкие фонетические классы, модель диктора, модели гауссовых смесей, сегментация речевого сигнала, кластеризация фонем.

In the article, principal technologies used in the creation of speaker identification systems and difficulties faced by their developers are considered. The structure of text-independent speaker identification using automatic segmentation of speech signal with simultaneous speaker-independent classification of segments is proposed. This approach improves accuracy of the speaker model and eliminates disagreement between training and recognizable context.
Key words: speaker identification, wide phonetic classes, speaker model, Gaussian mixture model, speech signal segmentation, clustering of phonemes.

У статті розглянуті основні технології, що використовуються при створенні систем ідентифікації диктора, і труднощі, з якими стикаються їх розробники. Запропоновано структуру системи текстонезалежної ідентифікації диктора, що використовує автоматичну дикторонезалежну сегментацію мовного сигналу з одночасною класифікацією сегментів. Такий підхід підвищує точність моделі диктора і нівелює суперечність між навчальним і розпізнавальним контекстом.
Ключові слова: ідентифікація диктора, широкі фонетичні класи, модель диктора, моделі гауссових сумішей, сегментація речового сигналу, кластеризація фонем.

Введение

Идентификация личности по голосу в настоящее время широко используется как отдельно, так и в совокупности с другими биометрическими показателями в системах безопасности, программных или аппаратных многопользовательских комплексах. Удобство и простота выполнения авторизации при помощи устной речи позволяет применять подобные системы удаленно (мобильная связь, сеть Интернет и т.д.).

Существует два основных типа систем голосовой биометрии: текстозависимые и текстонезависимые.

Текстозависимые применяются в системах контроля доступа: для верификации необходимо произнести парольную фразу, которая сравнивается с хранящимися в системе эталонами произнесения каждого зарегистрированного пользователя. Уязвимое место таких систем − получение несанкционированного доступа путем копирования парольной фразы современными средствами акустического прослушивания. Данный недостаток отсутствует в текстонезависимых системах.

Для верификации или аутентификации в текстонезависимых системах может использоваться практически любой фрагмент свободной звучащей речи достаточной длины, что делает их удобными с точки зрения пользователя. Такие системы незаменимы при решении полицейских задач: скрытая идентификация, криминалистическая идентификация, фоноучеты. Тем не менее, эта возможность усложняет реализацию текстонезависимых систем, понижает их надежность и скорость распознавания.

Идентификация по голосу основана на анализе уникальных характеристик речи, обусловленных анатомическими особенностями речевого тракта, а также приобретенными привычками произношения. На этапе извлечения признаков речевой сигнал сегментируется на короткие участки и на каждом участке вычисляется набор признаков. В качестве признаков для идентификации диктора в системах обоих типов используются различные параметры, учитывающие процессы как речеобразования (характеристики распределения частоты основного тона (ЧОТ), коэффициенты линейного предсказания, спектр Фурье), так и восприятия речи (вейвлет-спектр, мел-частотные кепстральные коэффициенты − MFCC), и их динамические характеристики. Все извлекаемые из аудиосигнала показатели не лишены недостатков: робастные параметры обладают слабыми идентификационными качествами и, наоборот, параметры, характеризующие диктора с высокой точностью, достаточно сильно чувствительны к различным факторам:

– нестабильность произнесения фразы диктором (темп, громкость произношения, физическое и эмоциональное состояние человека во время речевого акта);

– вид и уровень помех в акустическом и электронном канале связи, искажение речевого сигнала приемниками звука и реверберацией помещения.

Цель данной статьи – показ и предоставление полученных в настоящее время определенных результатов по исследованию эффективности систем идентификации голоса, которые показывают достаточно точную идентификацию и верификацию дикторов, когда эталон голоса клиента и его запрос поступают по одному и тому же каналу. Однако вопрос о создании особо точных систем идентификации по голосу, устойчивых к канальным искажениям, остается открытым. В связи с этим возникает ряд задач, таких, как исследование точности, робастности параметров и методов, используемых для идентификации по голосу, расширение поля признаков. Работы в этом направлении представляются более чем актуальными.

Современные технологии идентификации личности методами голосовой биометрии

Подавляющее большинство систем идентификации диктора имеют типовую укрупненную структуру, представленную на рис. 1. Перед выделением идентификационных характеристик также может происходить процедура компенсации канальных искажений. Принятие решения может происходить как с учетом множества признаков, так и на основе одного.

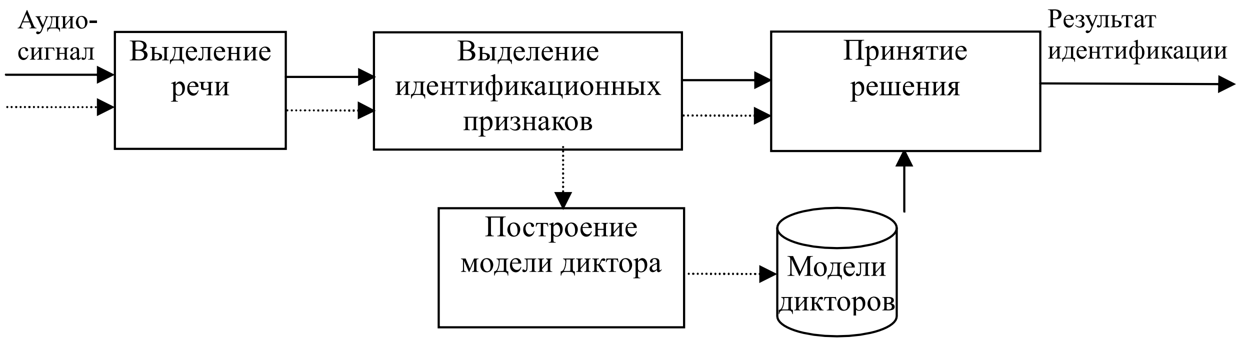


Рисунок 1 − Типовая структурная схема системы идентификации диктора в режимах обучения (пунктирная линия) и идентификации (сплошная линия)

Текстозависимые системы отличаются относительной простотой реализации, поскольку соотнесение полученных идентификационных характеристик с эталонными сводится к их тривиальному сравнению. Возможные ошибки вследствие разной скорости произнесения эталона парольной фразы и сравниваемого сигнала устраняются при помощи DTW [1]. Наиболее безопасными среди данных систем являются реализации с парами пользователь-пароль. В качестве характеристик диктора могут быть использованы любые акустические признаки речевого сигнала, наиболее используемыми являются форманты [2], как достаточно робастные идентификационные признаки.

В работе [3] приводится сравнительный анализ эффективности идентификации по фиксированным словам с помощью 14 наборов признаков. Среди которых лучшие результаты демонстрируют MFCC, непрерывное вейвлет-преобразование и коэффициенты отражения, получаемые с помощью кодирования с линейным предсказанием (КЛП). Текстозависимые системы не предъявляют особых требований к классификаторам, поэтому существуют реализации с различными типами классификаторов: линейными [4], на основе скрытых марковских моделей, меры Атала [5], нейросетей [3].

Событийнозависимые системы основаны на выявлении особенностей диктора в определенных фонемах и их последовательностях. В [6] идентификационные характеристики выделяются на участках глухих фрикативных звуков ([с], [ш]) с использованием нормированного количества импульсов равной длины. Сравнение с эталонами реализовано с помощью DTW. Данная методика позволила достичь 92% вероятности идентификации на множестве из 10 дикторов.

В работе [7] характеристики строятся на вокализованных отрезках речевого сигнала. Вектора признаков состоят из трех первых формант и трех антиформант, которые получаются из сглаженного спектра с помощью КЛП. Идентификация диктора, произведенная на основе нечеткой нейросети с обратным распространением ошибки, показала вероятность правильной идентификации диктора 93%, при обучении с помощью генетического алгоритма − 95%.

Однако, на практике данные системы применимы редко вследствие того, что они рассматривают только часть фонем, а следовательно, модель диктора сформирована неполно. Кроме того, необходимое количество искомых фонем может не содержаться в произвольной фразе, а добавление представительной базы фонем диктора требует длительного обучения системы. Поэтому методы анализа специфических фонемных классов чаще всего включаются в состав текстонезависимых систем.

Для проведения *текстонезависимой* идентификации существует два подхода. Первый заключается в том, что по акустическим признакам речевого сигнала для каждого диктора строятся статистические модели. Идентификация в данном случае представляет собой вычисление отклонения случайного вектора от модельных распределений и принятие решений происходит с заданным порогом допуска. Второй подход основан на создании в рамках одной системы гендеро- и канало-зависимых подсистем, функционирующих на отдельных наборах речевых признаков. Решение принимается в результате взвешенного голосования подсистем [8].

Примером текстонезависимой системы идентификации диктора с такой организацией может служить система, разработанная ООО «Центр речевых технологий» [8]. Система адаптирована для различных каналов связи (трех типов) и осуществляет гендерозависимую обработку входных данных, в качестве идентификационных признаков использует независимые линейно-частотные кепстральные коэффициенты и MFCC. Таким образом, формируется 6 отдельных подсистем, обученных отдельно на длительных аудиоданных. Идентификация основана на получении решения из обобщенных решений подсистем методом взвешенного голосования. При этом точность идентификации составляет 95%. Учитывая высокий уровень качества идентификации данной системы, следует отметить и ее недостатки: сложность построения моделей дикторов (большие объемы обучаемых выборок); устранение канальных искажений решено только дублированием моделей с использованием искажений (добавление нового типа искажения потребует увеличения количества подсистем); высокая степень гендерозависимости.

Для формирования модели диктора наиболее широкое применение получили:

– векторное квантование;

– гауссовы смеси;

– метод опорных векторов.

Идея векторного квантования заключается в следующем: при формировании эталона для конкретного диктора пространство признаков разбивается на непересекающиеся кластеры. Разбиение на кластеры считается индивидуальным, поэтому при идентификации говорящего по поступающему речевому сообщению распределение кластеров похоже на эталонное для зарегистрированного пользователя.

Результатом векторного квантования является кодовая книга. При ее формировании, как правило, используют процедуру кластеризации, также применяют методы нечеткой логики [9]. Улучшить результаты кластеризации можно с помощью метода максимизации правдоподобия.

Использование этого метода без предварительной кластеризации приводит к увеличению вычислительных операций.

При использовании модели гауссовых смесей, как и при векторном квантовании, предполагается, что акустическое пространство голоса диктора может быть характеризовано множеством акустических классов, отражающих некоторые особенности конфигурации его голосового тракта.

Модель гауссовых смесей описывает многомерное вероятностное распределение как взвешенную сумму множества более простых нормальных распределений, по-

лученных для каждого акустического класса, который представляется вектором математического ожидания, и ковариационной матрицей. Предполагая, что векторы признаков независимы друг от друга, плотность наблюдения векторов, образующих эти классы, можно считать смесью гауссовых распределений. В общем виде модель из *M* компонент представляется в виде

$$p\left(\overline{x} \mid \lambda\right) = \sum_{i=0}^{M} w_i\, p_i\left(\overline{x}\right), \quad \sum_{i=1}^{M} p_i = 1,$$

где $\overline{x}$ − *D*-мерный вектор признаков; w_i − вес *i*-го компонента модели, $p_i(\overline{x})$ − функция распределения *i*-го компонента модели. Каждый компонент описывается *D*-мерной гауссовой функцией распределения вида

$$p_i\left(\overline{x}\right) = \frac{1}{\left(2\Pi\right)^{D/2}\left(\Sigma_i\right)^{1/2}} \exp\left\{-\frac{1}{2}\left(\overline{x}-\overline{u}_i\right)^T\left(\Sigma_i\right)^{-1}\left(\overline{x}-\overline{u}_i\right)\right\},$$

где $\overline{u}_i$ − вектор математического ожидания и (Σ_i) − ковариационная матрица.

Полностью модель гауссовой смеси определяется векторами математического ожидания, ковариационными матрицами и весами смесей для каждого компонента модели. Эти параметры все вместе записываются в виде

$$\lambda = \left\{w_i, \overline{u}_i, \Sigma_i\right\}, i = 1,\ldots,M$$

Поскольку гауссовы смеси моделируют одну функцию плотности вероятности, то нет необходимости использовать полные ковариационные матрицы, даже если параметры вектора не являются полностью независимыми друг от друга.

Линейная комбинация диагональных ковариационных матриц способна моделировать корреляцию между элементами вектора наблюдений.

Эффект использования множества *M* ковариационных матриц может быть достигнут путем увеличения числа гауссовых компонент, использующих диагональные ковариационные матрицы [10].

Результат идентификации − модель диктора, которая имеет наибольшее значение апостериорной вероятности для произнесенной фразы, т.е.:

$$\sum_{i=0}^{M} p\left(\overline{x}_i \mid \lambda_k\right).$$

Этот критерий получил название «критерий максимального правдоподобия».

В отличие от векторного квантования, модель гауссовых смесей использует перекрывающиеся области в пространстве признаков.

В последнее время в качестве классификатора часто используется метод опорных векторов, строящий гиперплоскость, равноудаленную от выпуклых элементов противоположных классов.

Проблема линейно неразделимых классов решается вводом параметра допуска или применением ядрового преобразования, которое проецирует исходное пространство в пространство большей размерности.

Применение данного метода целесообразно в системах со значительным количеством классов и активно исследуется на предмет эффективного подбора ядер и оптимизации вычислений.

Основной целью исследований в области распознавания дикторов является создание алгоритмов, повышающих точность идентификации, сохраняющих при этом приемлемые показатели по вычислительной трудоемкости.

В данной работе предлагается подход к проектированию системы текстонезависимой идентификации говорящего, использующий дикторонезависимый блок. Это позволяет нивелировать разногласие между обучающим и распознаваемым контекстом.

Описание структуры проектируемой системы

При разработке структуры системы идентификации мы исходили из предположения, что множество классов, характеризующих акустическое пространство голоса диктора, описывает определенные фонетические события − звуки различных широких фонетических классов (ШФК) как гласные, фрикативные и т.д.

Предлагаемая в данной статье структура идентификации диктора призвана снизить влияние канальных искажений и увеличить качество идентификации за счет применения различных дикторонезависимых признаков. С целью разбиения произвольной речи на участки, принадлежащие различным ШФК, и создания для каждого диктора множества моделей для каждого фонетического элемента.

Структурная схема проектируемой системы текстонезависимой идентификации диктора приведена на рис. 2.

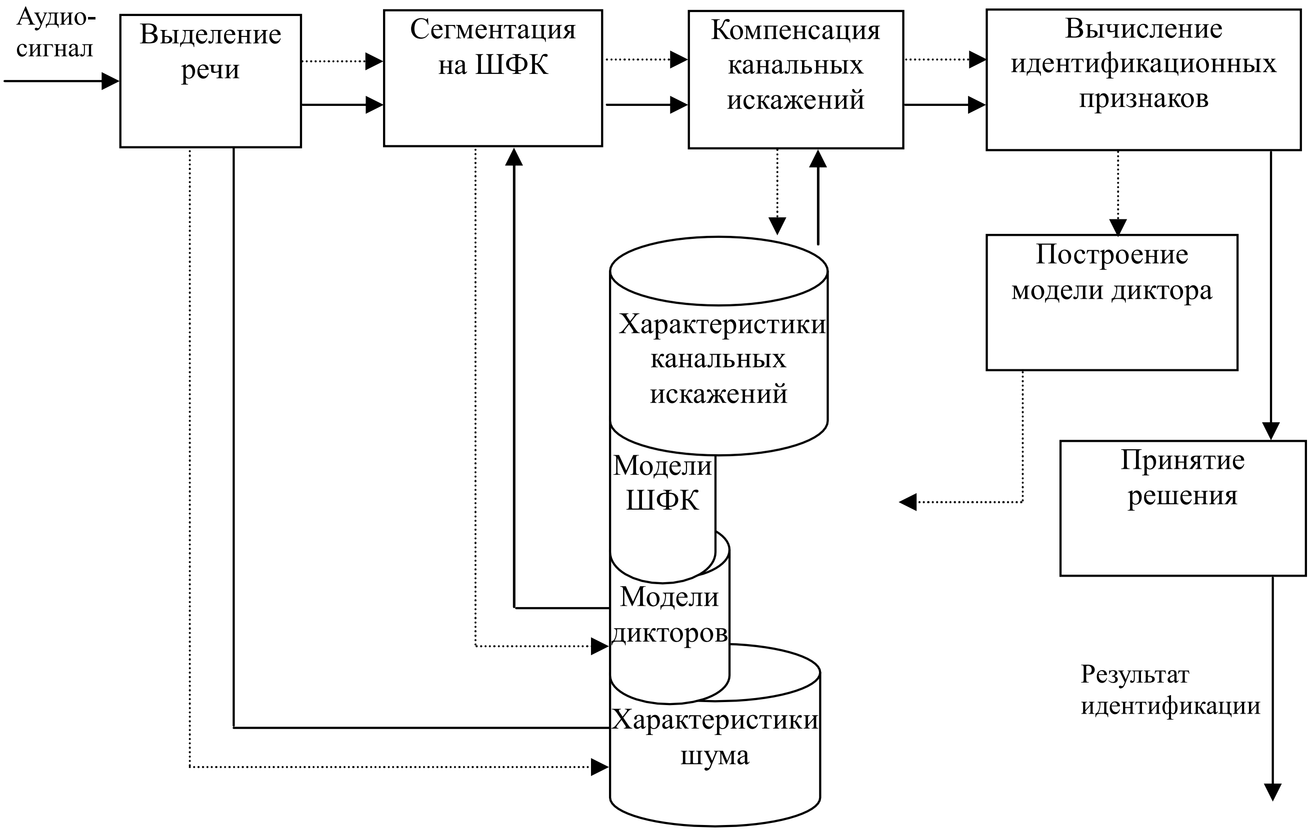


Рисунок 2 − Структурная схема текстонезависимой системы идентификации диктора в режимах обучения (пунктирная линия) и идентификации (сплошная линия)

Теперь о блоках системы более подробно.

Выделение речи из аудиосигнала проводится, исходя из условия, что начало аудиосигнала (0,5 сек) является участком шума. Для определения границ речи предполагается использовать хорошо зарекомендовавший себя метод, изложенный в [11].

Сегментация на широкие фонетические классы (ШФК) должна быть основана на дикторонезависимых характеристиках фонем. Для этой цели были выбраны MFCC, поскольку эти признаки небольшим набором коэффициентов (чаще всего − 13) информативно описывают акустические характеристики фонем.

Для проведения процедуры сегментации по обучающей выборке, полученной по речевым фрагментам нескольких дикторов, было сформировано пространство признаков и разбито на ШФК. Обучение выполнялось последовательно в два этапа:

1) кластеризация в рамках каждого ШФК (количество кластеров зависит от состава фонетического класса);

2) создание модели каждого ШФК на основе гауссовых смесей.

Для кластеризации был применен метод К-средних с итеративным добавлением центроидов (делением кластера с максимальным радиусом на два). Начальный центроид располагается в центре выборки, а в качестве критерия эффективности описания выборки применен ICL-BIC без использования штрафа на число компонент [12].

Условие, когда кластеризация считается завершенной при ухудшении данного критерия, показало высокую скорость сходимости и достаточное качество кластеризации.

Для уточнения положения центроидов использовался метод максимизации правдоподобия.

Результаты кластеризации каждого ШФК легли в основу его модели, которая создавалась с помощью гауссовых смесей размерностью 10.

По сформированным моделям выполнялась автоматическая сегментация тестовых речевых сигналов с одновременной классификацией их фреймов по критерию максимального правдоподобия, описанному выше.

Пример автоматической сегментации речевого фрагмента продемонстрирован на рис. 3.

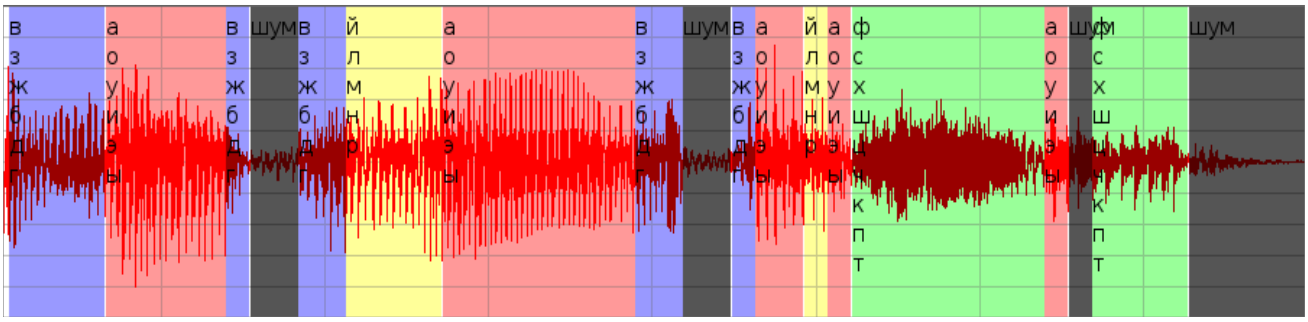


Рисунок 3 − Пример автоматической сегментации аудиосигнала

Одной из остро стоящих проблем дикторонезависимой сегментации является выбор состава ШФК. Согласно классификации звуков русской речи по их образованию можно выделить фонетические группы, представленные в табл. 1.

Был проведен ряд исследований для изучения влияния состава ШФК на результаты автоматической сегментации. Для построения моделей ШФК была проведена ручная сегментация на фонемы аудиозаписей речи 2 дикторов мужского и 2 − женского пола. Запись производилась с использованием динамического микрофона в незашумленной обстановке с частотой дискретизации 44,1 кГц и глубиной квантования 16 бит. Сформированное пространство признаков состояло из 3000 векторов. Эксперименты проводились как с исходным сигналом, так и с сигналом после предварительной обработки.

Таблица 1 − Классификация звуков русской речи

<table>
<tr><th>Обозначение</th><th>Состав</th><th colspan="5">Название</th></tr>
<tr><td>G_Sh</td><td>[ф], [с], [х], [ш], [ф’], [с’], [х’], [ш’]</td><td>щелевые</td><td rowspan="3">глухие</td><td rowspan="5">шумные</td><td rowspan="8">согласные</td><td rowspan="3">невокализованные</td></tr>
<tr><td>G_SSh</td><td>[ц], [ч]</td><td>смычно-щелевые/ аффрикаты</td></tr>
<tr><td>G_S</td><td>[к], [т], [п], [к’], [т’], [п’]</td><td>смычные</td></tr>
<tr><td>Z_Sh</td><td>[в], [з], [ж], [в’], [з’], [ж’]</td><td>щелевые</td><td rowspan="2">звонкие</td><td rowspan="6">вокализованные</td></tr>
<tr><td>Z_S</td><td>[б], [д], [г], [б’], [д’], [г’]</td><td>смычные</td></tr>
<tr><td>S_Sh</td><td>[й], [л], [л’]</td><td>щелевые</td><td colspan="2" rowspan="3">сонорные</td></tr>
<tr><td>S_S</td><td>[м], [н], [м’], [н’]</td><td>смычные</td></tr>
<tr><td>S_D</td><td>[р], [р’]</td><td>дрожащие</td></tr>
<tr><td>V</td><td>[и], [э], [о], [у], [а], [ы]</td><td colspan="4">гласные</td></tr>
</table>

Предварительная обработка состояла в удалении ЧОТ из исходного сигнала. Данный прием использовался для снижения зависимости параметров как от постоянной составляющей сигнала, так и от особенностей ЧОТ дикторов, что необходимо в контексте разделения на ШФК. ЧОТ вычислялась автокореляционным методом без дополнительной обработки. Удаление ЧОТ из сигнала производилось путем его обработки режекторным узкополосным фильтром заданной частоты.

Эффективность сегментации и классификации полученных сегментов по ШФК с различным составом приведена в табл. 2. Ошибка сегментации выражена отношением количества ошибочно определенных сегментов к общему количеству сегментов речевого сигнала в аудиозаписи.

Таблица 2 − Эффективность классификации на различные ШФК

Состав ШФК	Ошибка классификации	
	Исходный сигнал	Вычитание ЧОТ
{Не речь (NV)}, {вокализованные}, {невокализованные}	6,5%	3,3%
{NV}, {V+S_*}, {G_*+Z_*}	8,2%	7%
{NV}, {V}, {S_*}, {G_*+Z_*}	11,2%	8,5%
{NV}, {V}, {S_*}, {G_*}, {Z_S}, {Z_Sh}	18,6%	18%
{NV}, {V}, {S_*}, {G_S}, {G_Sh}, {G_SSh}, {Z_S}, {Z_Sh}	24%	22,5%

В ходе анализа полученных результатов было установлено следующее:

1) значения MFCC, вычисленные по реализациям звуков, произнесенных различными дикторами, значительно близки у глухих щелевых и смычно-щелевых фонем, поэтому целесообразно объединять эти классы звуков в один;

2) при автоматической сегментации достаточно часто наблюдался пропуск границ смычных фонем, что может объясняться их непродолжительностью звучания и влиянием на значения их признаков следующей гласной либо сонорной фонемы;

3) после предварительной обработки речевого сигнала классификация его сегментов показала лучшие результаты, однако эффективность значительно возрастает с уменьшением числа фонетических классов, где влияние ОТ вносит значительные коррективы в модель вокализованных фонем;

4) наиболее часто ошибки классификации возникали на участках, содержащих межфонемный переход, что объясняется влиянием соседних фонем на значения признаков.

Для *компенсации канальных искажений* планируется создать базу характеристик каналов, полученных по записям многих типов микрофонов. Для компенсации определяется тип микрофона, проводится логарифмирование спектра входного сигнала, что переводит влияние канала из мультипликативной помехи в аддитивную и позволяет использовать методы кепстрального вычитания.

Пространство признаков, в котором принимается решение о личности диктора, должно формироваться с учетом всех факторов процесса речеобразования: голосового источника, резонансных частот речевого тракта и их затуханий, а также динамикой управления артикуляцией. Поэтому при разработке блока *вычисления идентификационных признаков,* кроме широко используемых в современных системах идентификации по голосу линейно-частотных кепстральных коэффициентов и MFCC, планируется рассмотреть следующие параметры:

1) голосового источника − средняя частота основного тона, контур частоты основного тона, флюктуации частоты основного тона и форма импульса возбуждения;

2) спектральные характеристики речевого тракта − огибающая спектра, его средний наклон, формантные частоты и ширина их полос, параметры огибающей спектра невокализованных;

3) просодические характеристики, описывающие систему управления артикуляцией − динамика ЧОТ, длительность фонетических сегментов.

Голос диктора описывается множеством моделей, полученных по разным ШФК. В блоке *построения модели диктора* планируется реализовать несколько методов: гауссовы смеси, векторное квантование, метод опорных векторов.

Каждый классификатор обладает определенными преимуществами и недостатками, и по-разному реагирует на различие в условиях обучения и распознавания, а также на особенности голоса разных дикторов. Поэтому целесообразно использовать решения разных классификаторов, чтобы достичь минимально возможной ошибки распознавания. Учесть качество каждого классификатора возможно при принятии решения как взвешенной по их оценкам суммы решений. Это позволяет делать бустинг – метод усиления простых классификаторов, основанный на комбинировании примитивных «слабых» в один «сильный». В блоке *принятия решений* планируется реализовать наиболее известный алгоритм бустинга AdaBoost [13]. Он строит сильный алгоритм машинного обучения по набору слабых алгоритмов машинного обучения путем многократного прохождения по обучающей выборке и увеличения веса примеров, на которых слабые алгоритмы дают большую ошибку обучения.

Выводы

В данной статье сделан аналитический обзор современных технологий идентификации личности по голосу, разработана структура системы текстонезависимой идентификации, использующая модели ШФК. Анализ полученных результатов позволил сделать следующие выводы.

1 Несмотря на множество методов обработки речевого сигнала и идентификации диктора, они все чувствительны к качеству передачи речевого сигнала через каналы связи и вариативности произношения диктора.

2 Предложен подход к проектированию системы текстонезависимой идентификации говорящего, использующий дикторонезависимый блок, формирующий модели ШФК. Это позволяет нивелировать разногласие между обучающим и распознаваемым контекстом. Кроме того, для повышения робастности процесса идентификации предлагается в структуру системы включить блок компенсации канальных искажений с использованием соответствующей базы характеристик искажений.

3 Проведено исследование качества автоматической сегментации и классификации речевого сигнала на ШФК различного состава. При наиболее оптимальном с позиции разделимости составе ШФК, ошибка классификации для сигнала без предварительной обработки составила 11%, после обработки – 8%. Сложность для классификации представляют участки, содержащие межфонемный переход, для сегментации – короткие смычные фонемы.

Представляется возможным повысить эффективность автоматической сегментации речевого сигнала и классификации полученных сегментов за счет:

– увеличения точности и робастности методов вычисления ЧОТ путем частичного приглушения пиков гармоник, кратных ЧОТ, и определения оптимальной ширины полосы затухания фильтра, применяемого для удаления ЧОТ из речевого сигнала с целью уменьшения дикторозависимости процесса сегментации и классификации сегментов речевого сигнала;

– устранения пропуска границ коротких по времени фонем путем анализа речевого сигнала с регулируемым перекрытием окна;

– выбора параметров, формирующих вектор признаков голосовой модели диктора, по каждому ШФК с учетом особенностей составляющих его фонем.

Развитием данной работы для построения робастной текстонезависимой системы идентификации диктора могут стать следующие направления исследований:

– исследование робастности и гендерозависимости предложенного метода сегментации с одновременной классификацией полученных сегментов;

– исследование характеристик различных канальных искажений, построения их представительной базы данных для компенсации динамических помех в речевом сигнале;

– исследование возможности улучшения точности идентификации за счет введения дополнительных классификаторов (векторное квантование, машины опорных векторов) и построения на их основе сильного классификатора с помощью алгоритма бустинга.

Литература

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Винцюк Т.К. – К. : Наук. думка, 1987. – 261 с.
2. Ручай А.Н. К вопросу о законе распределения форманты, биометрической характеристики диктора / А.Н. Ручай // Проблемы теоретической и практической математики : тезисы 41-й Всероссийской молодежной конференции. − Екатеринбург : УрО РАН, 2010. − С. 401-407.
3. Федоров Е.Е. Методика идентификации диктора на основе модифицированной вероятностной нейронной сети / Е.Е. Федоров // Наукові праці ДонНТУ. Серія «Інформатика, кібернетика та обчислювальна техніка». − 2011. − № 13(185). − С. 186-191.
4. Венедиктова Е.В. Идентификация диктора по фиксированному набору частот с помощью линейного классификатора / Е.В. Венедиктова, Д.Н. Лавров // Математические структуры и моделирование. − 2008. − № 18. − С. 108-115 .
5. Атал Б.С. Автоматическое опознавание дикторов по голосам / Б.С. Атал // ТИИЭР. − 1976. − Т. 64, № 4. − С. 48-66.
6. Федоров Е.Е. Идентификация диктора на основе шипящих звуков / Е.Е. Федоров // Искусственный интеллект. − 2006. − № 4. − С. 197-206.
7. Федоров Е.Е. Методика идентификации водителя на основе формантного подхода и нечеткой нейросети / Федоров Е.Е., Ларин В.Ю., Слесорайтите Э. // Вісник Донецької академії автомобільного транспорту. − 2011. − № 4. − С. 35-43.

8. Матвеев Ю.Н. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 / Ю.Н. Матвеев, К.К. Симончик // Аннотация. – 5 с.
9. Любимов Н. Сравнение алгоритмов кластеризации в задаче идентификации диктора / Н. Любимов, Е. Михеев, А.С. Лукин. // Труды 13-й международной конференции «Цифровая обработка сигналов и её применение» (DSPA2011). − М. : 2011. − Т. 1. − С. 204-207.
10. Benesty J. Springer Handbook of Speech Processing / Benesty J., Sondhi M.M., Huang Y. – Springer-Verlag, 2008. – P 3.1, 7.1, 7.2.
11. Ермоленко Т.В. Классификация фреймов речевого сигнала в задачах дикторонезависимого распознавания речи / Т.В. Ермоленко, А.В. Жук // Искусственный интеллект. – 2011. – № 4. – С. 87-95.
12. Сорокин В.Н. Верификация диктора по спектрально-временным параметрам речевого сигнала / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – Т. 10, № 2. – С. 87-104.
13. Freund Y. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting / Y. Freund, R.E. Schapire // Journal of Computer and System Sciences. – 1997. – V. 55. – P. 119-139.

Literatura

1. Vintsyuk T.K. Analiz, raspoznavanie i interpretatsiya rechevykh signalov. K.: Nauk. Dumka. 1987. 261 s.
2. Ruchai A.N. Problemy tyeoreticheskoi i prakticheskoi matematiki: Tezisy 41-i Vserossiiskoi molodezhnoi konferentsii. Yekaterinburg: UrO RAN. 2010. S. 401-407.
3. Fedorov E.E. Naukovi pratsi DonNTU Seriya “Informatyka, kibernetyka ta obchyslyuvalna tekhnika”. 2011. № 13(185). S. 186-191.
4. Venediktova E.V. Matematicheskie struktury i modelirovanie. 2008. № 18. S. 108-115.
5. Atal B.S. TIIER. 1976. T. 64. № 4. S. 48-66.
6. Fedorov E.E. Iskusstvennyi intellekt. 2006. №4. S. 197-206.
7. Fedorov E.E. Visnyk Donetskoi akademii avtomobil’nogo transportu. 2011. № 4. S. 35-43.
8. Matvyeev U.N. Sistema identifikatsii diktorov po golosu dlya konkursa NIST SRE 2010. Annotatsiya. 5 s.
9. Lyubimov N. Trudy 13-i mezhdunarodnoi konferentsii “Tsifrovaya obrabotka signalov i ejo primenenie” (DSPA2011). M.: 2011. T. 1. S. 204-207.
10. Benesty J. Springer Handbook of Speech Processing. Springer-Verlag. 2008. P. 3.1, 7.1, 7.2.
11. Yermolenko T.V. Iskusstvennyi intellekt. 2011. № 4. S. 87-95.
12. Sorokin V.N. Informatsionnye protsessy. T. 10. № 2. S. 87-104.
13. Freund Y. A Journal of Computer and System Sciences. 1997. V. 55. P. 119-139.

RESUME

M.S. Klymenko

Development of Structure for Text-independent Speaker Identification System

In the article, methods used in speaker identification systems, main classes of voice biometrics, and difficulties faced by their developers are analyzed. After analysis of the described methods, the structure of text independent speaker identification system with addition of the channel distortion database and block of automatic speaker-independent segmentation of the speech signal into sections containing different phonemes of broad phonetic classes (BPCs) with simultaneous classification is proposed. Maintain database of channel distortion model allows storing compact speaker model and eliminating the use of sub-systems, adapted to the different audio channels. Using classification of speaker-independent segments will neutralize the difference between training and recognizable context and allow creating a set of speaker models obtained by different BPCs. This can significantly improve the efficiency of identification. BPCs model for segmentation of the available speech database formed with use of Gaussian mixture. Mel-frequency cepstral coefficients are used as acoustic features of phonemes. Formed on the model, automatic segmentation of test speech signals is performed. Simultaneous classification of their frames by maximum likelihood is also performed. The investigation of the dependence of quality speaker-independent segmentation on the composition of BPCs is performed. It shows improvement of the segmentation quality by reducing the number of classes. In addition, efficiency of the classification increases for the pre-processed signal. It consists in the removal of fundamental frequency. Pretreatment was applied to reduce dependence of phonemes on the speaker voice. The best result is shown by the phonetic classification of four classes with signal preprocessing, the error is 8%.

Статья поступила в редакцию 05.07.2012.